

University of Wollongong

Research Online

National Institute for Applied Statistics
Research Australia Working Paper Series

Faculty of Engineering and Information
Sciences

2015

The SAR model for very large datasets: A reduced rank approach

Sandy Burden

University of Wollongong

Noel Cressie

University of Wollongong

David Steel

University of Wollongong

Follow this and additional works at: <https://ro.uow.edu.au/niasrawp>

Recommended Citation

Burden, Sandy; Cressie, Noel; and Steel, David, The SAR model for very large datasets: A reduced rank approach, National Institute for Applied Statistics Research Australia, University of Wollongong, Working Paper 09-15, 2015, 22.

<https://ro.uow.edu.au/niasrawp/30>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

The SAR model for very large datasets: A reduced rank approach

Abstract

The SAR model is widely used in spatial econometrics to model Gaussian processes on a discrete spatial lattice, but for large datasets fitting it becomes computationally prohibitive, and hence its usefulness can be limited. A computationally efficient spatial model is the Spatial Random Effects (SRE) model, and in this article we calibrate it to the SAR model of interest using a generalisation of the Moran operator that allows for heteroskedasticity and an asymmetric SAR-spatial-dependence matrix. In general, spatial data have a measurement-error component, which we model, and we use restricted maximum likelihood to estimate the SRE-model covariance parameters; its required computational time is only the order of the size of the dataset. Our implementation is demonstrated using mean usual weekly income data from the 2011 Australian Census.

NIASRA

NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

09-15

The SAR model for very large datasets: A reduced rank approach

Sandy Burden, Noel Cressie and David Steel

*Copyright © 2015 by the National Institute for Applied Statistics Research Australia, UOW.
Work in progress, no part of this paper may be reproduced without permission from the Institute.*

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845.
Email: anica@uow.edu.au

The SAR model for very large datasets: A reduced rank approach

Sandy Burden ^{1,*}, Noel Cressie ¹ and David Steel ¹

¹ National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, NSW 2522, Australia

* Author to whom correspondence should be addressed; sburden@uow.edu.au, +61 (0)2 4221 5084.

Version: 20 April 2015

Abstract: The SAR model is widely used in spatial econometrics to model Gaussian processes on a discrete spatial lattice, but for large datasets fitting it becomes computationally prohibitive, and hence its usefulness can be limited. A computationally efficient spatial model is the Spatial Random Effects (SRE) model, and in this article we calibrate it to the SAR model of interest using a generalisation of the Moran operator that allows for heteroskedasticity and an asymmetric SAR-spatial-dependence matrix. In general, spatial data have a measurement-error component, which we model, and we use restricted maximum likelihood to estimate the SRE-model covariance parameters; its required computational time is only the order of the size of the dataset. Our implementation is demonstrated using mean usual weekly income data from the 2011 Australian Census.

Keywords: Asymmetric Spatial-Dependence Matrix; Australian Census; Heteroskedasticity; Moran Operator; Spatial Autoregressive Model; Spatial Basis Functions; Spatial Random Effects Model

1. Introduction

The spatial autoregressive (SAR) model, originally proposed by Whittle [1] [see also, 2], is widely used in spatial econometrics to model Gaussian processes on a discrete spatial lattice (which may be irregular). It has an autoregressive structure that models spatial dependence of the attributes through a precision matrix that is typically a function of the proximity between elements of the lattice. This spatial-dependence matrix is generally sparse. An important application of the SAR model is when

the lattice is made up of n spatially contiguous areas, such that for two areas that are “far apart” the corresponding entry in the spatial-dependence matrix is zero.

Despite its popularity, the SAR model has some limitations for the analysis of large datasets, because maximum likelihood estimation of its parameters is not scalable in the size of the spatial dataset. This limits the size of the problems for which SAR models can be effectively used. Algorithmic approaches to reduce the computational burden include sparse-matrix techniques [3,4] and approximate methods such as Taylor series and Chebyshev approximations [4,5].

Rather than modelling spatial dependence through a precision matrix that is a function of the spatial proximity of areas, it could be modelled directly through a covariance matrix; a reduced-rank covariance model could then be used to lower the order of computations [e.g., 6] to $O(n)$. Reduced-rank approaches include latent-factor-analysis models [7]; predictive-process models for Gaussian data [8,9] and non-Gaussian data [10]; spatial random effects (SRE) models [11,12]; and spatial generalised linear mixed models (SGLMMs) [13–15]. The advantage of the SRE approach is a substantial improvement in computational efficiency for a broad class of flexible, spatial, non-stationary covariance functions [12,16]. The spatial dependence is captured through the choice of basis functions and through an $r \times r$ covariance matrix of the coefficients of the basis functions. By reducing the rank of the spatial covariance function from n to r through the use of r spatial basis functions, the SRE model has a computational cost of $O(nr^2)$, which is $O(n)$ for $r \ll n$ [12]. Our challenge in this paper is to calibrate and fit a computationally efficient SRE model so that its parameters are directly interpretable in terms of the familiar SAR model.

An important aspect of model definition is specifying appropriate basis functions. Several classes of basis functions have previously been used for SRE models, including empirical orthogonal functions (EOFs) [summarised in, e.g., 6,17], Fourier basis functions [18], smoothing splines [6], radial basis functions [e.g., 19, and 20, pages 186–187], and multi-resolutional basis functions (i.e., basis functions obtained from functions defined at multiple resolutions) such as bisquare spatial basis functions [12, 21] and wavelets [16,22], that capture spatial dependence at multiple scales. Selecting an appropriate class of basis functions is generally problem-dependent, although the use of the conditional Akaike Information Criterion and the generalised-degrees-of-freedom criterion have been considered [23]. The spatial basis functions selected for the model need not be orthogonal, but they are recommended to be multi-resolutional and fast to evaluate [12]. More recently, a spectral decomposition of the matrix in the numerator of Moran’s I statistic, called the Moran operator [14], has been used to identify a set of orthogonal eigenvectors associated with dominant patterns of positive spatial dependence [e.g., 24–26]. These eigenvectors have been used to obtain spatial basis functions for SGLMMs [27], and substantial computational efficiency has been achieved for both SGLMMs [14] and SRE models [28] by including only the first $r \ll n$ basis functions in the model.

Moran’s I can be motivated from a simple SAR model with homoskedastic errors and symmetric spatial-dependence matrix given by nearest-neighbour adjacencies. In this article, we consider a heteroskedastic SAR model with asymmetric spatial-dependence matrix and define a Generalised Moran operator (GMO) whose eigenvectors define our spatial basis functions. In particular, we propose to model spatial data for small areas using an SRE model with these spatial basis functions *calibrated* from the proposed SAR model. The SAR-calibrated SRE model, for which computational efficiencies are

readily available, is well suited to very large datasets, and its parameters are interpretable in terms of the original SAR parameters.

The plan for the rest of the article is as follows. Section 2 outlines the main features of SAR and SRE models. A motivation is given in Section 3 for spatial basis functions obtained from a Generalised Moran operator. Dimension-reduction using these basis functions is developed in Section 4. Parameter estimation under the SAR-calibrated SRE model is presented in Section 5. An example of their use for Australian Census data is presented in Section 6, and the paper concludes with a discussion in Section 7.

2. SAR Model and SRE Model Specifications

Consider a set of n geographic areas $\{A_i : i = 1, \dots, n\}$ in a spatial domain, and identify the location of a well defined feature in each area, denoted $\mathbf{s}_1, \dots, \mathbf{s}_n$, respectively. Define an irregular, finite spatial lattice $D_s \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ over the region of interest, $\bigcup_{i=1}^n A_i$, and consider the attributes $\{Y(\mathbf{s}_i) : i = 1, \dots, n\}$ associated with the n areas.

We are interested in making inference on a spatial process $Y(\cdot)$, which we henceforth assume is Gaussian. However, $Y(\cdot)$ is a latent process: Rather than observing $\{Y(\mathbf{s}_i) : i = 1, \dots, n\}$, we observe data $\{Z(\mathbf{s}_i) : i = 1, \dots, n\}$, which offer an imperfect view of the latent process due to limitations in the measurement process (including averaging, the use of proxy measures rather than direct measurement, and observation error).

In this section, we specify hierarchical models that allow us to make inference on the latent process $\mathbf{Y} \equiv (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))'$ using the observed data $\mathbf{Z} \equiv (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$. We first define a SAR model for \mathbf{Y} and then a measurement-error model for \mathbf{Z} given \mathbf{Y} . Subsequently, we consider an SRE model for \mathbf{Y} and discuss the implications of equating the SAR and SRE models' respective means, variances, and covariances.

2.1. SAR Model

A SAR model for \mathbf{Y} over the lattice D_s , which includes a spatially varying mean, is defined as:

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \sum_{j=1}^n b_{ij}(Y(\mathbf{s}_j) - \mu(\mathbf{s}_j)) + v(\mathbf{s}_i), \quad i = 1, \dots, n, \quad (1)$$

where $\boldsymbol{\mu} \equiv (\mu(\mathbf{s}_1), \dots, \mu(\mathbf{s}_n))'$ is an n -dimensional vector representing the mean of \mathbf{Y} , and b_{ij} is a measure of the spatial dependence between the locations \mathbf{s}_i and \mathbf{s}_j with the convention that $b_{ii} = 0$; $i = 1, \dots, n$. We define the SAR's $n \times n$ spatial-dependence matrix as $\mathbf{B} \equiv \{b_{ij} : i, j = 1, \dots, n\}$. In general, $b_{ij} \neq b_{ji}$; that is, \mathbf{B} is an asymmetric square matrix. The elements of the error vector $\mathbf{v} \equiv (v(\mathbf{s}_1), \dots, v(\mathbf{s}_n))'$ are independent Gaussian random variables with mean zero and variance $\sigma_{v_i}^2$, $i = 1, \dots, n$. Their joint distribution can be written as $\mathbf{v} \sim \text{Gau}(\mathbf{0}, \boldsymbol{\Sigma}_v)$, where $\boldsymbol{\Sigma}_v \equiv \text{diag}(\sigma_{v_1}^2, \dots, \sigma_{v_n}^2)$. For many applications, $\boldsymbol{\Sigma}_v$ is parameterised as $\sigma_{v_i}^2 = \sigma_v^2 V_v(\mathbf{s}_i)$, where $\sigma_v^2 > 0$, and $V_v(\cdot) > 0$ is a known function of location. We write $\boldsymbol{\Sigma}_v \equiv \sigma_v^2 \mathbf{V}_v$, where the diagonal matrix $\mathbf{V}_v \equiv \text{diag}(V_v(\mathbf{s}_1), \dots, V_v(\mathbf{s}_n))$ may be the identity or may model heteroskedasticity (e.g., the inverse of the population size in each area; see Section 6).

Here we model the mean as a linear function of p known covariates; that is, $\boldsymbol{\mu} \equiv \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X} \equiv (\mathbf{X}(\mathbf{s}_1), \dots, \mathbf{X}(\mathbf{s}_n))'$, for $\mathbf{X}(\mathbf{s}_i) \equiv (X_1(\mathbf{s}_i), \dots, X_p(\mathbf{s}_i))'$, and $\boldsymbol{\beta} \equiv (\beta_1, \dots, \beta_p)'$ is the associated p -dimensional vector of regression coefficients. In this case, (1) can be written equivalently as,

$$\mathbf{Y} \equiv \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \mathbf{B})^{-1}\mathbf{v}. \quad (2)$$

The spatial-dependence matrix \mathbf{B} has zeros down the diagonal, generally it is asymmetric, and it is often defined in terms of the proximity of the lattice elements. For (2) to be well defined, we consider only those \mathbf{B} for which $(\mathbf{I} - \mathbf{B})$ is invertible. Generally, \mathbf{B} (and hence $(\mathbf{I} - \mathbf{B})$) is sparse. In many applications, \mathbf{B} is parameterised as $\rho\mathbf{H}$, where $\mathbf{H} \equiv \{h_{ij} : i, j = 1, \dots, n\}$ is a known, in general asymmetric, $n \times n$ spatial-weights matrix with zeros down the diagonal, and ρ is a spatial-dependence parameter. The case of $\rho = 0$ corresponds to “no spatial dependence,” since $\mathbf{B} = \rho\mathbf{H} = \mathbf{0}$.

Henceforth, we parameterise our model as $\mathbf{B} \equiv \rho\mathbf{H}$, and $\boldsymbol{\Sigma}_v \equiv \sigma_v^2\mathbf{V}_v$, where \mathbf{H} and \mathbf{V}_v are known. Consequently, there are two covariance parameters, ρ and σ_v^2 . From (2), the mean vector and the covariance matrix of \mathbf{Y} are, respectively, given by,

$$\begin{aligned} E_{SAR}(\mathbf{Y}) &\equiv \mathbf{X}\boldsymbol{\beta} \\ \text{cov}_{SAR}(\mathbf{Y}) &\equiv \sigma_v^2(\mathbf{I} - \rho\mathbf{H})^{-1}\mathbf{V}_v(\mathbf{I} - \rho\mathbf{H}')^{-1}. \end{aligned} \quad (3)$$

A statistical model for the observations, \mathbf{Z} , is defined in terms of the latent process, \mathbf{Y} , by assuming a conditional Gaussian distribution for the observed data. The resulting data model can be written in terms of additive measurement error, as follows:

$$Z(\mathbf{s}_i) = Y(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), \quad i = 1, \dots, n. \quad (4)$$

The term $\varepsilon(\cdot)$ in (4) represents an uncorrelated Gaussian process with mean zero and individual variances given by $\text{var}(\varepsilon(\mathbf{s}_i)) = \sigma_\varepsilon^2 V_\varepsilon(\mathbf{s}_i)$, for $i = 1, \dots, n$, where $\sigma_\varepsilon^2 > 0$, and $V_\varepsilon(\cdot) > 0$ is a known function of location; $\varepsilon(\cdot)$ is also independent of $Y(\cdot)$. Hence,

$$\mathbf{Z}|\mathbf{Y} \sim \text{ind. Gau}(\mathbf{Y}, \sigma_\varepsilon^2\mathbf{V}_\varepsilon),$$

where “Gau” is the Gaussian (or normal) distribution, and $\mathbf{V}_\varepsilon \equiv \text{diag}(V_\varepsilon(\mathbf{s}_1), \dots, V_\varepsilon(\mathbf{s}_n))$ may be the identity matrix, or it may represent heterogeneous conditions associated with the measurement process. Together, (2) and (4) yield the following model for \mathbf{Z} ,

$$\mathbf{Z} \equiv \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{H})^{-1}\mathbf{v} + \boldsymbol{\varepsilon}, \quad (5)$$

where $\boldsymbol{\varepsilon} \equiv (\varepsilon(\mathbf{s}_1), \dots, \varepsilon(\mathbf{s}_n))' \sim \text{Gau}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{V}_\varepsilon)$, independent of $\mathbf{v} \sim \text{Gau}(\mathbf{0}, \sigma_v^2\mathbf{V}_v)$. Traditional spatial econometrics has fitted a SAR model directly to the spatial data, \mathbf{Z} , as follows ([e.g., 2]): $\mathbf{Z} \equiv \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{H}\mathbf{Z} + \mathbf{v}$. We believe that (5) presents the three components of variability, namely trend, SAR spatial dependence, and measurement error, in an unambiguous and interpretable manner.

Motivated by Besag *et al.* [29], we have assumed a SAR model for the spatial errors, $\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$, to avoid confounding $\rho\mathbf{H}$ with $\boldsymbol{\beta}$. Furthermore, to recognise that measurement is never perfect, we have added measurement error to the latent process \mathbf{Y} . Details regarding these models and their estimation can be found in, for example, Anselin [2], Cliff and Ord [30], Cressie [31], Le Sage [32], Anselin [33], and LeSage and Pace [34].

2.2. SRE Model

The SRE model is a reduced-rank spatial statistical model that can be written as follows:

$$Y(\mathbf{s}_i) = \mathbf{X}(\mathbf{s}_i)' \boldsymbol{\beta} + \mathbf{S}(\mathbf{s}_i)' \boldsymbol{\eta} + \xi(\mathbf{s}_i), \quad i = 1, \dots, n, \quad (6)$$

where $\mathbf{S}(\cdot) \equiv (S_1(\cdot), \dots, S_r(\cdot))'$ is a known vector of $r \ll n$ spatial basis functions, and $\boldsymbol{\eta} \equiv (\eta_1, \dots, \eta_r)'$ is its corresponding r -dimensional vector of random effects. The random effects are assumed to have a multivariate Gaussian distribution with mean $\mathbf{0}$ and $r \times r$ covariance matrix \mathbf{K} , so that $\boldsymbol{\eta} \sim \text{Gau}(\mathbf{0}, \mathbf{K})$. The last term in (6), $\xi(\cdot)$, represents fine-scale variation, independent of \mathbf{Y} , which is assumed here to follow,

$$\xi(\mathbf{s}_i) \sim \text{ind. Gau}(0, \sigma_\xi^2 V_\xi(\mathbf{s}_i)), \quad i = 1, \dots, n,$$

where $\sigma_\xi^2 > 0$, and $V_\xi(\cdot) > 0$ is a known function of location. Hence, $\boldsymbol{\xi} \sim \text{Gau}(\mathbf{0}, \sigma_\xi^2 \mathbf{V}_\xi)$, where $\boldsymbol{\xi} \equiv (\xi(\mathbf{s}_1), \dots, \xi(\mathbf{s}_n))'$, and the diagonal matrix $\mathbf{V}_\xi \equiv \text{diag}(V_\xi(\mathbf{s}_1), \dots, V_\xi(\mathbf{s}_n))$ may be the identity matrix or may represent heterogeneity, for example through (population size)⁻¹ in each area; see Section 6. Using vector notation, (6) can also be written as,

$$\mathbf{Y} \equiv \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\eta} + \boldsymbol{\xi}, \quad (7)$$

where $\mathbf{S} \equiv (\mathbf{S}(\mathbf{s}_1), \dots, \mathbf{S}(\mathbf{s}_n))'$. The mean vector and the covariance matrix of \mathbf{Y} for the SRE model are, respectively, given by,

$$\begin{aligned} E_{SRE}(\mathbf{Y}) &\equiv \mathbf{X}\boldsymbol{\beta} \\ \text{cov}_{SRE}(\mathbf{Y}) &\equiv \mathbf{S}\mathbf{K}\mathbf{S}' + \sigma_\xi^2 \mathbf{V}_\xi. \end{aligned} \quad (8)$$

As was done for the SAR model, we define \mathbf{Z} in terms of the latent process \mathbf{Y} through the data model, (4). Together, (7) and (4) yield the following model for \mathbf{Z} :

$$\mathbf{Z} \equiv \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\eta} + \boldsymbol{\xi} + \boldsymbol{\varepsilon}. \quad (9)$$

We calibrate the specification of \mathbf{S} , $\text{cov}(\boldsymbol{\eta})$, and $\text{cov}(\boldsymbol{\xi})$ in the SRE model to the spatial dependence in the SAR model, by initially matching their first two moments. Now, the mean of the SRE model (7) is the same as the mean of the SAR model (2). Then, after setting the measurement-error terms in (5) and (9) to be the same, we set $\text{cov}_{SRE}(\mathbf{Y}) \equiv \text{cov}_{SAR}(\mathbf{Y})$ [e.g., 35]. That is, we set

$$\mathbf{S}\mathbf{K}\mathbf{S}' + \sigma_\xi^2 \mathbf{V}_\xi = \sigma_v^2 (\mathbf{I} - \rho \mathbf{H})^{-1} \mathbf{V}_v (\mathbf{I} - \rho \mathbf{H}')^{-1}. \quad (10)$$

Notice that in the SRE model (left-hand side), the contribution to spatial dependence is from the reduced-rank matrix $\mathbf{S}\mathbf{K}\mathbf{S}'$, whilst in the SAR model (right-hand side), the contribution to spatial dependence comes from the full-rank matrix, $\mathbf{I} - \rho \mathbf{H}$. When $\mathbf{S}\mathbf{K}\mathbf{S}' = \mathbf{0} = \rho \mathbf{H}$, we respectively obtain an SRE model and an SAR model with no spatial dependence. In this case, the SRE model and the SAR model are equivalent if

$$\sigma_\xi^2 \mathbf{V}_\xi = \sigma_v^2 \mathbf{V}_v. \quad (11)$$

This says that heteroskedasticity in the SRE's fine-scale-variation component $\boldsymbol{\xi}$ should reflect the heteroskedasticity in the SAR model's \mathbf{v} , when there is no spatial dependence.

When spatial dependence is present in the SAR model, the spatial-dependence component of the SRE model should, according to (10), be equal to:

$$\sigma_v^2(\mathbf{I} - \rho\mathbf{H})^{-1}\mathbf{V}_v(\mathbf{I} - \rho\mathbf{H}')^{-1} - \sigma_\xi^2\mathbf{V}_\xi. \quad (12)$$

Critically, (12) should also be positive-definite with rank $r \ll n$. Ensuring that these restrictions hold leads to our new methodology, which is presented in Section 4.

3. Motivation for the Spatial Basis Functions

We use this section to motivate the construction of our SAR-calibrated spatial basis functions for use in the SRE model. The motivation is clearest for the case where $\sigma_\varepsilon^2 = 0$, and hence in this section we temporarily assume $\mathbf{Z} = \mathbf{Y}$. The SAR-model parameters are $(\boldsymbol{\beta}, \rho, \sigma_v^2)$. Minus twice the log-likelihood function for the SAR model is

$$\begin{aligned} -2\mathcal{L}(\boldsymbol{\beta}, \rho, \sigma_v^2; \mathbf{Y}) &= n \log(2\pi) - \log \left(\left| \sigma_v^{-2}(\mathbf{I} - \rho\mathbf{H}')\mathbf{V}_v^{-1}(\mathbf{I} - \rho\mathbf{H}) \right| \right) + \sigma_v^{-2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{I} - \rho\mathbf{H}')\mathbf{V}_v^{-1}(\mathbf{I} - \rho\mathbf{H})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= n \log(2\pi) - \log \left(\left| \sigma_v^{-2}(\mathbf{V}_v^{-1/2} - \rho\mathbf{H}'\mathbf{V}_v^{-1/2})(\mathbf{V}_v^{-1/2} - \mathbf{V}_v^{-1/2}\rho\mathbf{H}) \right| \right) + \\ &\quad \sigma_v^{-2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{V}_v^{-1/2} - \rho\mathbf{H}'\mathbf{V}_v^{-1/2})(\mathbf{V}_v^{-1/2} - \mathbf{V}_v^{-1/2}\rho\mathbf{H})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned} \quad (13)$$

Consider the transformation $\tilde{\mathbf{Y}} = \mathbf{V}_v^{-1/2}\mathbf{Y}$. The spatial process $\tilde{\mathbf{Y}}$ is distributed as $\tilde{\mathbf{Y}} \sim \text{Gau}(\tilde{\mathbf{X}}\boldsymbol{\beta}, \sigma_v^2(\mathbf{I} - \rho\tilde{\mathbf{H}})^{-1}(\mathbf{I} - \rho\tilde{\mathbf{H}}')^{-1})$, where $\tilde{\mathbf{X}} = \mathbf{V}_v^{-1/2}\mathbf{X}$, and $\tilde{\mathbf{H}} = \mathbf{V}_v^{-1/2}\mathbf{H}\mathbf{V}_v^{1/2}$ [e.g., see 36, for details]. Hence (13) can be written as,

$$\begin{aligned} -2\mathcal{L}(\boldsymbol{\beta}, \rho, \sigma_v^2; \mathbf{Y}) &= n \log(2\pi) - \log \left(\left| \sigma_v^{-2}(\mathbf{V}_v^{-1/2} - \mathbf{V}_v^{-1/2}\rho\tilde{\mathbf{H}}')(\mathbf{V}_v^{-1/2} - \rho\tilde{\mathbf{H}}\mathbf{V}_v^{-1/2}) \right| \right) + \\ &\quad \sigma_v^{-2}(\mathbf{V}_v^{1/2}\tilde{\mathbf{Y}} - \mathbf{V}_v^{1/2}\tilde{\mathbf{X}}\boldsymbol{\beta})'(\mathbf{V}_v^{-1/2} - \mathbf{V}_v^{-1/2}\rho\tilde{\mathbf{H}}')(\mathbf{V}_v^{-1/2} - \rho\tilde{\mathbf{H}}\mathbf{V}_v^{-1/2})(\mathbf{V}_v^{1/2}\tilde{\mathbf{Y}} - \mathbf{V}_v^{1/2}\tilde{\mathbf{X}}\boldsymbol{\beta}) \\ &= n \log(2\pi) - 2 \log |\mathbf{V}_v^{-1/2}| - \log \left(\left| \sigma_v^{-2}(\mathbf{I} - \rho\tilde{\mathbf{H}}')(\mathbf{I} - \rho\tilde{\mathbf{H}}) \right| \right) + \sigma_v^{-2}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\mathbf{I} - \rho\tilde{\mathbf{H}}')(\mathbf{I} - \rho\tilde{\mathbf{H}})(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) \\ &= -2\mathcal{L}(\boldsymbol{\beta}, \rho, \sigma_v^2; \tilde{\mathbf{Y}}) + \log |\mathbf{V}_v|. \end{aligned} \quad (14)$$

The SAR-model maximum likelihood estimates, $(\hat{\boldsymbol{\beta}}, \hat{\rho}, \hat{\sigma}_v^2)$, are those values that minimise $-2\mathcal{L}(\boldsymbol{\beta}, \rho, \sigma_v^2; \mathbf{Y})$, or equivalently they are those values that minimise $-2\mathcal{L}(\boldsymbol{\beta}, \rho, \sigma_v^2; \tilde{\mathbf{Y}})$.

Maximum likelihood estimates for the covariance parameters ρ and σ_v^2 are well known to exhibit negative bias [e.g., 37]. To reduce the bias, estimation commonly proceeds using restricted (or residual) maximum likelihood (REML) [e.g., see 38, for a summary of its use for spatial data]. REML estimation is carried out by performing maximum likelihood estimation on $\mathbf{M}\tilde{\mathbf{Y}}$, where \mathbf{M} is chosen such that $\mathbf{E}(\mathbf{M}\tilde{\mathbf{Y}}) = \mathbf{0}$, $\text{rank}(\mathbf{M}) = n - p$, and p is the rank of the matrix of covariates, $\tilde{\mathbf{X}}$. Suppose we choose $\mathbf{M} = \tilde{\mathbf{P}}^\perp \equiv \mathbf{I} - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'$. Note that $\tilde{\mathbf{P}}^\perp$ is symmetric (i.e., $(\tilde{\mathbf{P}}^\perp)' = \tilde{\mathbf{P}}^\perp$) and idempotent (i.e., $(\tilde{\mathbf{P}}^\perp)^2 = \tilde{\mathbf{P}}^\perp$). Effectively, we are substituting an ordinary-least-squares estimator of $\boldsymbol{\beta}$ (based on $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{X}}$) into (14), and then we minimise this empirical version of the likelihood to obtain REML estimates. The

“data” are now $\tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}}$, with $E(\tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}}) = \mathbf{0}$ and $\text{cov}(\tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}}) = \tilde{\mathbf{P}}^\perp \text{cov}(\tilde{\mathbf{Y}}) \tilde{\mathbf{P}}^\perp$. Hence we minimise the restricted likelihood,

$$\begin{aligned} -2\mathcal{L}(\boldsymbol{\beta}, \rho, \sigma_v^2; \tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}}) &= n \log(2\pi) - \log \left(\left| \sigma_v^{-2} \tilde{\mathbf{P}}^\perp (\mathbf{I} - \rho \tilde{\mathbf{H}}') (\mathbf{I} - \rho \tilde{\mathbf{H}}) \tilde{\mathbf{P}}^\perp \right| \right) + \\ &\quad \sigma_v^{-2} (\tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}} - \tilde{\mathbf{P}}^\perp \tilde{\mathbf{X}} \boldsymbol{\beta})' \tilde{\mathbf{P}}^\perp (\mathbf{I} - \rho \tilde{\mathbf{H}}') (\mathbf{I} - \rho \tilde{\mathbf{H}}) \tilde{\mathbf{P}}^\perp (\tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}} - \tilde{\mathbf{P}}^\perp \tilde{\mathbf{X}} \boldsymbol{\beta}) \\ &= n \log(2\pi) - \log \left(\left| \sigma_v^{-2} \tilde{\mathbf{P}}^\perp (\mathbf{I} - \rho \tilde{\mathbf{H}}') (\mathbf{I} - \rho \tilde{\mathbf{H}}) \tilde{\mathbf{P}}^\perp \right| \right) + \sigma_v^{-2} \tilde{\mathbf{Y}}' \tilde{\mathbf{P}}^\perp (\mathbf{I} - \rho \tilde{\mathbf{H}}') (\mathbf{I} - \rho \tilde{\mathbf{H}}) \tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}}, \end{aligned} \quad (15)$$

resulting in the REML estimates. From (15), these estimates depend critically on the non-negative-definite matrix,

$$\sigma_v^{-2} \tilde{\mathbf{P}}^\perp (\mathbf{I} - \rho \tilde{\mathbf{H}}') (\mathbf{I} - \rho \tilde{\mathbf{H}}) \tilde{\mathbf{P}}^\perp. \quad (16)$$

Expanding (16) gives the equivalent expression,

$$\sigma_v^{-2} \left(\tilde{\mathbf{P}}^\perp \tilde{\mathbf{P}}^\perp - 2\rho \frac{\tilde{\mathbf{P}}^\perp (\tilde{\mathbf{H}}' + \tilde{\mathbf{H}}) \tilde{\mathbf{P}}^\perp}{2} + \rho^2 \tilde{\mathbf{P}}^\perp \tilde{\mathbf{H}}' \tilde{\mathbf{H}} \tilde{\mathbf{P}}^\perp \right). \quad (17)$$

In (17), the dominant term in ρ is proportional to

$$\tilde{\mathbf{P}}^\perp \left\{ \frac{\tilde{\mathbf{H}}' + \tilde{\mathbf{H}}}{2} \right\} \tilde{\mathbf{P}}^\perp. \quad (18)$$

Notice that Moran's I statistic for data $\tilde{\mathbf{Y}}$ and covariates $\tilde{\mathbf{X}}$ is,

$$\mathbf{I}(\mathbf{A}) = \frac{n}{\mathbf{1}' \mathbf{A} \mathbf{1}} \frac{\tilde{\mathbf{Y}}' \tilde{\mathbf{P}}^\perp \mathbf{A} \tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}}}{\tilde{\mathbf{Y}}' \tilde{\mathbf{P}}^\perp \tilde{\mathbf{Y}}}, \quad (19)$$

where \mathbf{A} is the adjacency matrix whose entries are 1 (when two entries are “adjacent”) and 0 (otherwise); see [24]. When an asymmetric matrix \mathbf{B} is used in place of \mathbf{A} , Tiefelsdorf [39, page 29] has suggested that \mathbf{B} be “symmetrised” by replacing \mathbf{A} in (19) with the symmetric matrix $(\mathbf{B}' + \mathbf{B})/2$.

We have seen that (18) plays an important role in the restricted likelihood (15), and it takes the form of the numerator of Tiefelsdorf's modification to the Moran statistic for an asymmetric spatial-dependence matrix. Hence, we use it to define a generalised Moran operator. Recall that \mathbf{V}_v captures heterogeneity and \mathbf{H} captures asymmetric spatial dependence in the SAR model. We can write $\tilde{\mathbf{P}}^\perp \equiv \mathbf{I} - \mathbf{V}_v^{-1/2} \mathbf{X} (\mathbf{X}' \mathbf{V}_v^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}_v^{-1/2}$ and, since $\tilde{\mathbf{H}} = \mathbf{V}_v^{-1/2} \mathbf{H} \mathbf{V}_v^{1/2}$,

$$\begin{aligned} \tilde{\mathbf{P}}^\perp (\tilde{\mathbf{H}}' + \tilde{\mathbf{H}}) \tilde{\mathbf{P}}^\perp &= \tilde{\mathbf{P}}^\perp \mathbf{V}_v^{1/2} (\mathbf{H}' \mathbf{V}_v^{-1} + \mathbf{V}_v^{-1} \mathbf{H}) \mathbf{V}_v^{1/2} \tilde{\mathbf{P}}^\perp \\ &= \mathbf{Q}(\mathbf{V}_v) (\mathbf{H}' \mathbf{V}_v^{-1} + \mathbf{V}_v^{-1} \mathbf{H}) \mathbf{Q}(\mathbf{V}_v)', \end{aligned}$$

where

$$\mathbf{Q}(\mathbf{V}_v) \equiv \mathbf{V}_v^{1/2} - \mathbf{V}_v^{-1/2} \mathbf{X} (\mathbf{X}' \mathbf{V}_v^{-1} \mathbf{X})^{-1} \mathbf{X}'. \quad (20)$$

Hence, we define the Generalized Moran operator (GMO),

$$\mathbf{G}(\mathbf{X}, \mathbf{H}, \mathbf{V}_v) \equiv \mathbf{Q}(\mathbf{V}_v) \left\{ \frac{\mathbf{H}' \mathbf{V}_v^{-1} + \mathbf{V}_v^{-1} \mathbf{H}}{2} \right\} \mathbf{Q}(\mathbf{V}_v)', \quad (21)$$

where $\mathbf{Q}(\mathbf{V}_v)$ is defined by (20).

Notice that for $\mathbf{V}_v = \mathbf{I}$ and $\mathbf{H} = \mathbf{A}$ (the symmetric adjacency matrix) we obtain $\mathbf{Q}(\mathbf{I}) = \mathbf{P}^\perp \equiv \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and hence (21) is equal to $\mathbf{P}^\perp \mathbf{A} \mathbf{P}^\perp$. This is the Moran operator given by Griffith [24] and Hughes and Haran [14]. Spectral decomposition of the Moran operator has been shown to yield natural spatial basis functions for spatial modelling [e.g., see 40, 41, 24, 27 and, more recently, 14, 28], and we propose the same for the GMO given by (21). For spatial generalised linear models where heteroskedasticity appears through a mean-variance relationship, Reich *et al.* [27] suggest modifying the Moran operator using weights obtained from iteratively reweighted least squares. This bears a formal resemblance to $\tilde{\mathbf{P}}^\perp$ and lends support to our proposed generalisation. Notice that the GMO allows for both heteroskedasticity (through \mathbf{V}_v) and asymmetry (through \mathbf{H}) simultaneously.

4. Specification of the SAR-calibrated SRE model

In this section, we define a SAR-calibrated SRE model for the latent process \mathbf{Y} , using the SRE model that is defined in Section 2.2. Our approach is to use $r \ll n$ eigenvectors from the spectral decomposition of $\mathbf{G}(\mathbf{X}, \mathbf{H}, \mathbf{V}_v)$ given by (21), to define the r basis functions for the SRE model, (8). Now, the spectral decomposition of (21) yields,

$$\mathbf{G}(\mathbf{X}, \mathbf{H}, \mathbf{V}_v) = \tilde{\mathbf{P}}^\perp \left\{ \frac{\tilde{\mathbf{H}}' + \tilde{\mathbf{H}}}{2} \right\} \tilde{\mathbf{P}}^\perp \equiv \mathbf{U} \mathbf{\Omega} \mathbf{U}', \quad (22)$$

where the eigenvectors \mathbf{U} of $\mathbf{G}(\mathbf{X}, \mathbf{H}, \mathbf{V}_v)$ satisfy $\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}$; and $\mathbf{\Omega} \equiv \text{diag}(\omega_1, \dots, \omega_n)$ is an $n \times n$ diagonal matrix with elements in the diagonal matrix ordered so that $|\omega_1| \geq |\omega_2| \geq \dots \geq |\omega_n| \geq 0$, for $\omega_1, \dots, \omega_n$, the eigenvalues of $\mathbf{G}(\mathbf{X}, \mathbf{H}, \mathbf{V}_v)$. Define the $n \times r$ matrix $\mathbf{U}_1 \equiv (\mathbf{u}_1, \dots, \mathbf{u}_r)$, which corresponds to the r largest *absolute* eigenvalues of \mathbf{G} , namely the diagonal elements in $\mathbf{\Omega}_1 \equiv \text{diag}(\omega_1, \dots, \omega_r)$. Similarly, \mathbf{U}_2 and $\mathbf{\Omega}_2$, correspond to the $n - r$ smallest absolute eigenvectors. Then,

$$\mathbf{U} \mathbf{\Omega} \mathbf{U}' = (\mathbf{U}_1, \mathbf{U}_2) \begin{pmatrix} \mathbf{\Omega}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_2 \end{pmatrix} \begin{pmatrix} \mathbf{U}_1' \\ \mathbf{U}_2' \end{pmatrix} = \mathbf{U}_1 \mathbf{\Omega}_1 \mathbf{U}_1' + \mathbf{U}_2 \mathbf{\Omega}_2 \mathbf{U}_2'. \quad (23)$$

Notice that eigenvectors corresponding to large *positive and negative* eigenvalues may be chosen; Hughes and Haran [14] only choose eigenvectors that correspond to large positive eigenvalues, which limits their approach to capturing only positive spatial dependence. In the application given in Section 6, several large absolute eigenvectors that corresponded to negative dependence were chosen.

From (3), using a second-order Taylor-series expansion for ρ around 0, the SAR-model covariance matrix for $\tilde{\mathbf{Y}}$ is:

$$\sigma_v^2 (\mathbf{I} - \rho \tilde{\mathbf{H}})^{-1} (\mathbf{I} - \rho \tilde{\mathbf{H}}')^{-1} \simeq \sigma_v^2 \left\{ \left[\mathbf{I} + \rho (\tilde{\mathbf{H}}' + \tilde{\mathbf{H}}) + \rho^2 \left((\tilde{\mathbf{H}}')^2 + \tilde{\mathbf{H}}' \tilde{\mathbf{H}} + \tilde{\mathbf{H}}^2 \right) \right] \right\}. \quad (24)$$

Hence, a candidate for the $r \times r$ matrix \mathbf{K} in the SRE model (8), is obtained by pre- and post-multiplying (24) by $\tilde{\mathbf{U}}_1$, where $\tilde{\mathbf{U}}_1 \equiv \mathbf{V}_v^{-1/2} \mathbf{U}_1$:

$$\sigma_v^2 \left\{ \tilde{\mathbf{U}}_1' \left[\mathbf{I} + \rho (\tilde{\mathbf{H}}' + \tilde{\mathbf{H}}) + \rho^2 \left((\tilde{\mathbf{H}}')^2 + \tilde{\mathbf{H}}' \tilde{\mathbf{H}} + \tilde{\mathbf{H}}^2 \right) \right] \tilde{\mathbf{U}}_1 \right\}. \quad (25)$$

The candidate \mathbf{K} given by (25) is fast to evaluate for given values of ρ and σ_v^2 . However, while the $r \times r$ matrix in (25) is symmetric, it is not necessarily positive-definite. Higham [42] shows how to obtain the

nearest symmetric positive approximant to any real square matrix \mathbf{B} . Call this positive-definite matrix $\mathcal{A}^+(\mathbf{B})$, and hence define,

$$\mathbf{R}^+(\rho) \equiv \mathcal{A}^+ \left(\tilde{\mathbf{U}}_1' \left[\mathbf{I} + \rho(\tilde{\mathbf{H}}' + \tilde{\mathbf{H}}) + \rho^2 \left((\tilde{\mathbf{H}}')^2 + \tilde{\mathbf{H}}' \tilde{\mathbf{H}} + \tilde{\mathbf{H}}^2 \right) \right] \tilde{\mathbf{U}}_1 \right). \quad (26)$$

Then, since $\sigma_v^2 > 0$, the $r \times r$ matrix ($r \ll n$),

$$\mathbf{K}^+(\rho, \sigma_v^2) \equiv \sigma_v^2 \mathbf{R}^+(\rho), \quad (27)$$

is positive-definite. In conclusion, we shall use the $r \times r$ positive-definite matrix \mathbf{K}^+ given by (27) to define our SAR-calibrated SRE model. What remains in the calibration is specification of $\sigma_\xi^2 \mathbf{V}_\xi$ in (8). Since \mathbf{V}_ξ is assumed known, we only need to specify σ_ξ^2 , which will generally be a function of ρ and σ_v^2 . We use the *closeness of the inverses* of the SAR and SRE covariance matrices to achieve this, which is concomitant with the eigenvectors in \mathbf{U}_1 coming from large absolute eigenvalues.

Recall that the SAR model's inverse covariance matrix is easy to evaluate, as is the SRE model's inverse covariance matrix using the Sherman-Morrison-Woodbury identity [e.g., 43]: For generic matrices \mathbf{A} , \mathbf{B} , \mathbf{U} , and \mathbf{V} of appropriate order,

$$(\mathbf{A} + \mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{U} (\mathbf{B}^{-1} + \mathbf{VA}^{-1} \mathbf{U})^{-1} \mathbf{VA}^{-1},$$

where the inverses on the right-hand side are fast to evaluate. Consequently, for an SRE model for $\tilde{\mathbf{Y}}$ with the form of (8), and with $\tilde{\mathbf{S}} = \tilde{\mathbf{U}}_1$, diagonal $\tilde{\mathbf{V}}_\xi = \mathbf{V}_v^{-1/2} \mathbf{V}_\xi \mathbf{V}_v^{-1/2}$, and positive-definite $\mathbf{K} \equiv \mathbf{K}^+(\rho, \sigma_v^2) = \sigma_v^2 \mathbf{R}^+(\rho)$, the *inverse* covariance matrix is,

$$\begin{aligned} (\tilde{\mathbf{S}} \mathbf{K} \tilde{\mathbf{S}}' + \sigma_\xi^2 \tilde{\mathbf{V}}_\xi)^{-1} &= (\tilde{\mathbf{U}}_1 \sigma_v^2 \mathbf{R}^+(\rho) \tilde{\mathbf{U}}_1' + \sigma_\xi^2 \tilde{\mathbf{V}}_\xi)^{-1} \\ &= \sigma_v^{-2} \{ \zeta \tilde{\mathbf{V}}_\xi^{-1} - \zeta^2 \tilde{\mathbf{V}}_\xi^{-1} \tilde{\mathbf{U}}_1 \left(\mathbf{R}^+(\rho)^{-1} + \zeta \tilde{\mathbf{U}}_1' \tilde{\mathbf{V}}_\xi^{-1} \tilde{\mathbf{U}}_1 \right)^{-1} \tilde{\mathbf{U}}_1' \tilde{\mathbf{V}}_\xi^{-1} \} \\ &\equiv \sigma_v^{-2} \mathbf{J}(\rho, \zeta), \end{aligned} \quad (28)$$

where $\zeta \equiv \sigma_v^2 / \sigma_\xi^2$. Consider the Frobenius norm between the *inverse* of the SAR covariance matrix of $\tilde{\mathbf{Y}}$ and (28); this is proportional to

$$\|(\mathbf{I} - \rho \tilde{\mathbf{H}}')(\mathbf{I} - \rho \tilde{\mathbf{H}}) - \mathbf{J}(\rho, \zeta)\|_F^2, \quad (29)$$

as a function of ζ . The final step in the calibration is to minimise (29) with respect to $\zeta > 0$, for all ρ in its parameter space; call the minimised value, $\zeta(\rho)$.

Hence, the SAR-calibrated SRE covariance matrix of \mathbf{Y} is,

$$\begin{aligned} \Sigma(\rho, \sigma_v^2) &\equiv \mathbf{V}_v^{1/2} \text{cov}(\tilde{\mathbf{Y}}) \mathbf{V}_v^{1/2} \\ &= \sigma_v^2 \{ \mathbf{U}_1 \mathbf{R}^+(\rho) \mathbf{U}_1' + \zeta(\rho)^{-1} \mathbf{V}_\xi \}, \end{aligned} \quad (30)$$

where $\sigma_v^2 \mathbf{R}^+(\rho)$ is a positive-definite $r \times r$ matrix, $\zeta(\rho) > 0$, by construction, and recall that \mathbf{V}_ξ is known. Notice that all parameters in the SRE model's covariance matrix (30) are expressed as functions of the SAR model's parameters σ_v^2 and ρ . Also notice that, while we have made $\Sigma(\rho, \sigma_v^2)^{-1}$ close (in the Frobenius-norm sense) to $(\mathbf{I} - \rho \mathbf{H}') \mathbf{V}_v^{-1} (\mathbf{I} - \rho \mathbf{H})$, they are not equal; this difference is the *only* source

of approximation in this article. The dominant (positive and negative) spatial-dependencies in the SAR model are included in the calibrated model in a reduced-rank form, and residual variability is captured by setting $\sigma_\xi^2 = \sigma_v^2/\zeta$ (which is a function of ρ). In the next section, we fit the model,

$$\mathbf{Z} \sim \text{Gau}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\rho, \sigma_v^2) + \sigma_\varepsilon^2 \mathbf{V}_\varepsilon), \quad (31)$$

using likelihood-based methods, where $\boldsymbol{\Sigma}(\rho, \sigma_v^2)$ is given by (30).

5. Parameter Estimation

We use likelihood-based estimators of the parameters of the SAR-calibrated SRE model (31), because they have asymptotic optimality properties, and they are preserved under transformation. The SAR model's parameters are $\boldsymbol{\beta}$, ρ , and σ_v^2 , which appear in the SRE likelihood defined by (30) and (31); we assume that the measurement-error component, $\sigma_\varepsilon^2 \mathbf{V}_\varepsilon$, is known, which is often the case (Section 6.3).

We estimate the SAR covariance parameters ρ and σ_v^2 using the log-restricted-likelihood function given by:

$$\mathcal{L}_R(\rho, \sigma_v^2) = \text{constant} - \frac{1}{2}(\log |\boldsymbol{\Phi}| + \log |\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{X}| + \mathbf{Z}'\boldsymbol{\Pi}\mathbf{Z}), \quad (32)$$

where $\boldsymbol{\Phi} \equiv \sigma_v^2 \{ \mathbf{U}_1 \mathbf{R}^+(\rho) \mathbf{U}_1' + \zeta(\rho)^{-1} \mathbf{V}_\xi \} + \sigma_\varepsilon^2 \mathbf{V}_\varepsilon$, and $\boldsymbol{\Pi} \equiv \boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Phi}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Phi}^{-1}$. We solve for (ρ, σ_v^2) by minimising $-2\mathcal{L}_R(\rho, \sigma_v^2)$ in terms of ρ and $\sigma_v^2 \geq 0$ in two stages. First, for each value of ρ , we minimise $-2\mathcal{L}_R(\rho, \sigma_v^2)$ with respect to σ_v^2 , to obtain $\hat{\sigma}_v^2(\rho)$. Then we minimise the profile, $-2\mathcal{L}_R(\rho, \hat{\sigma}_v^2(\rho))$, with respect to ρ . The resulting $\hat{\rho}$ and $\hat{\sigma}_v^2 \equiv \hat{\sigma}_v^2(\hat{\rho})$, are the REML estimates. When making likelihood calculations, we use the Sherman-Morrison-Woodbury formulas [43] for both the inverse and the determinant of the SRE model's covariance function.

This estimation technique requires a starting value for σ_v^2 for each value of ρ ; we use method-of-moments estimation to obtain the starting values. Specifically, define $\mathbf{W} \equiv \mathbf{P}^\perp \mathbf{Z}$, and hence

$$\mathbf{W} \sim \text{Gau}(\mathbf{0}, \sigma_v^2 \mathbf{L}(\rho) + \mathbf{M}), \quad (33)$$

where $\mathbf{L}(\rho) \equiv \mathbf{P}^\perp \{ \mathbf{U}_1 \mathbf{R}^+(\rho) \mathbf{U}_1' + \zeta(\rho) \mathbf{V}_\xi \} \mathbf{P}^\perp$ and $\mathbf{M} \equiv \sigma_\varepsilon^2 \mathbf{P}^\perp \mathbf{V}_\varepsilon \mathbf{P}^\perp$. Consequently,

$$E(\mathbf{W}'\mathbf{N}\mathbf{W}) = \text{trace}(\mathbf{N}\{\sigma_v^2 \mathbf{L}(\rho) + \mathbf{M}\}),$$

where \mathbf{N} is any symmetric, $n \times n$ matrix. For example, in Section 6 we choose $\mathbf{N} = \text{diag}(n_1, \dots, n_m)$, where n_i , $i = 1, \dots, m$ is the number of people in area i . The resulting method-of-moments estimator is,

$$\sigma_{v,0}^2(\rho) = \frac{\mathbf{Z}'\mathbf{P}^\perp \mathbf{N} \mathbf{P}^\perp \mathbf{Z} - \text{trace}(\sigma_\varepsilon^2 \mathbf{N} \mathbf{P}^\perp \mathbf{V}_\varepsilon \mathbf{P}^\perp)}{\text{trace}(\mathbf{N}\{\mathbf{P}^\perp \mathbf{U}_1 \mathbf{R}^+(\rho) \mathbf{U}_1' \mathbf{P}^\perp + \zeta(\rho)^{-1} \mathbf{P}^\perp \mathbf{V}_\xi \mathbf{P}^\perp\})}, \quad (34)$$

which we use as a starting value for solving:

$$\hat{\sigma}_v(\rho)^2 = \arg \min_{\sigma_v(\rho)^2} \{-2\mathcal{L}_R(\rho, \sigma_v^2)\}, \quad (35)$$

for each ρ in its parameter space. The profile restricted log-likelihood for ρ is given by,

$$\mathcal{L}_R^*(\rho) \equiv \mathcal{L}_R(\rho, \hat{\sigma}_v^2(\rho)),$$

for ρ in its parameter space. The minimum of $-2\mathcal{L}_R^*(\rho)$ is found using a univariate grid-search algorithm.

A likelihood-based confidence interval (CI) for a model parameter, with a significance level of α , can be obtained by inverting the (restricted) likelihood-ratio test. Hence an approximate $100(1 - \alpha)\%$ CI for ρ is given by:

$$\{\rho : \mathcal{L}_R^*(\rho) > \mathcal{L}_R(\hat{\rho}, \hat{\sigma}_v^2(\hat{\rho})) - \chi_1^2(1 - \alpha)/2\}, \quad (36)$$

where $\chi_1^2(1 - \alpha)$ is the $(1 - \alpha)100$ percentile of the chi-squared distribution on one degree of freedom.

Finally, the empirical generalised-least-squares estimate for β is,

$$\hat{\beta} \equiv (\mathbf{X}'\Phi(\hat{\rho}, \hat{\sigma}_v^2(\hat{\rho}))^{-1}\mathbf{X})^{-1}\mathbf{X}'\Phi(\hat{\rho}, \hat{\sigma}_v^2(\hat{\rho}))^{-1}\mathbf{Z}, \quad (37)$$

and an estimate of its covariance matrix is $(\mathbf{X}'\Phi(\hat{\rho}, \hat{\sigma}_v^2(\hat{\rho}))^{-1}\mathbf{X})^{-1}$.

6. Fitting a Spatial Model to Mean Usual Weekly Income

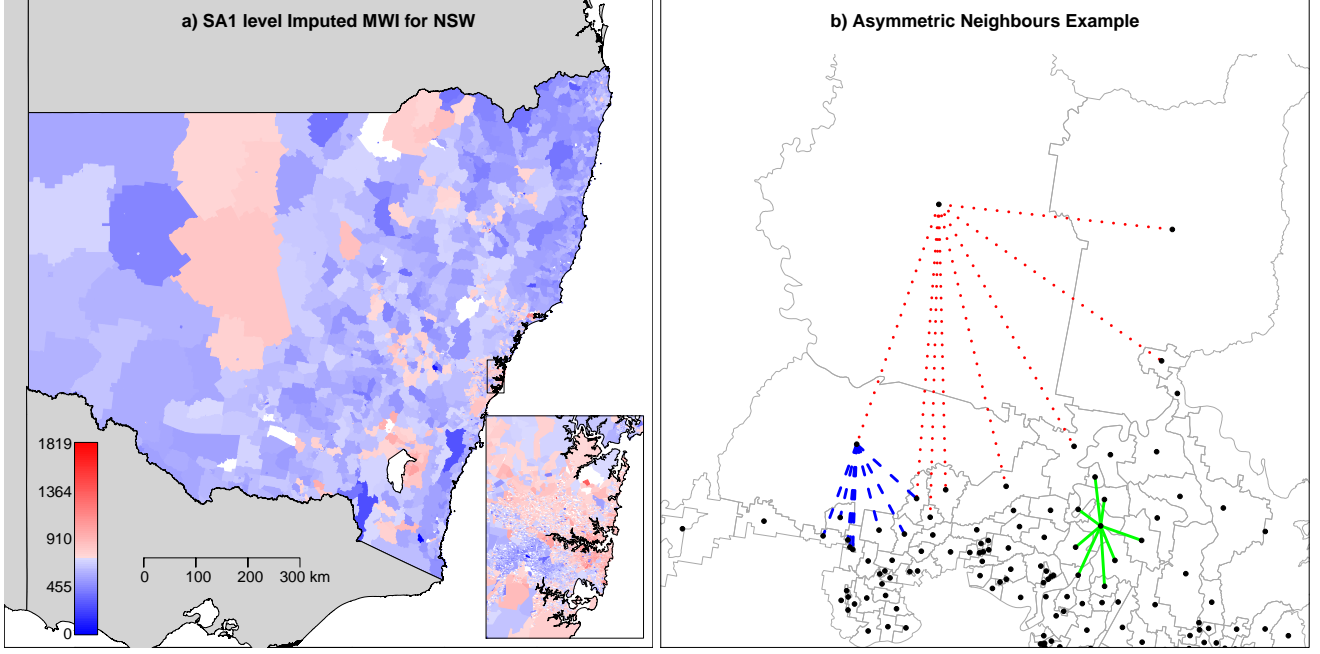
We illustrate the methodology developed in Sections 2 – 5 using Australian Census data [44] from the 2011 Census conducted by the Australian Bureau of Statistics (ABS). Our interest lies in modelling the spatial distribution of mean usual weekly income (MWI) for small geographic areas in the state of New South Wales (NSW), Australia, and then fitting the spatial model by estimating its parameters. In the following sections, we describe the data, the fitting of a SAR-calibrated SRE model, our assumptions regarding measurement error, and the statistical inferences that follow.

6.1. Imputed MWI

In NSW, 2011 Census summaries are available for four nested sets of geographic areas, called Statistical Areas (SA), where the most disaggregated level consists of SA1 areas. We fit a spatial model to MWI for the 16850 SA1 areas (Figure 1) with population > 100 persons, which is 97% of the total SA1 areas; populations ≤ 100 are too granular for the statistical analysis we shall undertake. SA1 areas in NSW exhibit a wide range in both population and area, since the state's population is concentrated along the coastline (in the east of the state). For the SA1 areas we consider, their populations range from 100 to 13000 (with a median of 395).

In the Census, each person ≥ 15 years specifies their usual mean weekly income by selecting from 11 possible income ranges $\{\leq \$0, \$1-\$199, \$200-\$299, \$300-\$399, \$400-\$599, \$600-\$799, \$800-\$999, \$1000-\$1249, \$1250-\$1499, \$1500-\$1999, \geq \$2000\}$. For the $i = 1, \dots, 16850$ SA1 areas, the Census data are the number of persons, n_{ij} , in the j th income range. Thus the 2011 Census does not directly measure MWI for SA1s; denote the true MWI process as $Y(\cdot)$. However, in the 2011-12 Australian Survey of Income and Housing (SIH) [45], the mean usual weekly income of a large sample of individuals was measured, but with limited spatial information. Hence, we use the SIH microdata to impute an MWI value, $\bar{Z}(s_i)$, for 2011 Census SA1 area $i = 1, \dots, 16850$, where the “bar” on $\bar{Z}(\cdot)$ is added to emphasise that MWI refers to *mean* income. In this analysis, we use only the SIH data and Census counts from the $J = 9$ Census income ranges corresponding to $\$0 < \text{MWI} < \2000 , which includes most (approximately 94%) of the incomes in the Census. Negative, zero, and very large incomes are set aside. These incomes, along with SA1 areas with populations less than 100, need special attention that is beyond the goal of this section.

Figure 1. a) Choropleth map of imputed MWI ($\$0 < \text{MWI} < \2000) for SA1 areas in NSW, Australia; the resolution around Sydney is enhanced. b) An excerpt of SA1 areas from a), showing how asymmetry can occur in the spatial-dependence matrix. The lines join the centroids of neighbouring areas, which are obtained using a K-nearest-neighbour algorithm, where here $K = 8$.



To impute the MWI values, we first stratify the Australia-wide SIH data using the Census income ranges, resulting in \tilde{n}_j SIH records $\{\tilde{Z}_{jl} : l = 1, \dots, \tilde{n}_j\}$ for the $j = 1, \dots, 9$ income ranges. The SIH data include weights that are used to account for unequal selection probabilities when calculating population estimates. We use them to impute n_{ij} MWI values, $\{Z_{ijk} : k = 1, \dots, n_{ij}\}$, by sampling n_{ij} SIH records, with replacement, from $\{\tilde{Z}_{jl} : l = 1, \dots, \tilde{n}_j\}$, for each of $i = 1, \dots, 16850$, $j = 1, \dots, 9$. As mentioned above, the MWI values in the samples are selected using the same probability distribution that was used to select individuals in the SIH survey. We average the imputed values for each area, to obtain an imputed MWI value:

$$\bar{Z}(s_i) \equiv \frac{\sum_{j=1}^9 \sum_{k=1}^{n_{ij}} Z_{ijk}}{n_i}, \text{ for } i = 1, \dots, 16850, \quad (38)$$

where $n_i \equiv \sum_{j=1}^9 n_{ij}$. Hence, the spatial data that we shall analyse in this section are,

$$\mathbf{Z} \equiv (\bar{Z}(s_1), \dots, \bar{Z}(s_{16850}))'. \quad (39)$$

6.2. A Simple Heteroskedastic Model for MWI

A basic heteroskedastic model for MWI was explored and found to fit reasonably well: Assume that for each area i and stratum j , income values are a random sample from a distribution with mean μ_j and variance σ_j^2 . From (38) and for given $\{n_{ij}\}$,

$$E(\bar{Z}(s_i)) = \frac{\sum_{j=1}^9 n_{ij} \mu_j}{n_i}, \quad \text{and} \quad \text{var}(\bar{Z}(s_i)) = \frac{\sum_{j=1}^9 (n_{ij}/n_i) \sigma_j^2}{n_i}. \quad (40)$$

Contrast this with an overly simple model that assumes a constant mean and variance for the elements of \mathbf{Z} .

In Figure 2, the left panel gives a histogram and Q-Q plot of unstandardised $\{\bar{Z}(\mathbf{s}_i)\}$ motivated by the overly simple model that assumes constant mean and variance over the SA1 areas. The right panel gives a histogram and Q-Q plot of the standardised values from the basic heteroskedastic model:

$$\frac{\bar{Z}(\mathbf{s}_i) - E(\bar{Z}(\mathbf{s}_i))}{\text{var}(\bar{Z}(\mathbf{s}_i))^{1/2}}, \text{ for } i = 1, \dots, 16850, \quad (41)$$

where $\{\mu_j\}$ and $\{\sigma_j^2\}$ in (40) are estimated straightforwardly from the SIH data. The fit of the histogram in the right panel to a $\text{Gau}(0, 1)$ distribution is striking and indicates that any spatial model we fit should aim to capture the same types of heteroskedastic behaviour given by (40). Of course $\{n_{ij}\}$ is an important part of the randomness in \mathbf{Z} defined by (38) and (39), and in a formal (spatial) model we cannot have them describing the first two moments (as we do informally in (40)).

To capture the heteroskedasticity in (40) and illustrated in Figure 2, recall from (5) that $\Sigma_v = \sigma_v^2 \mathbf{V}_v$. We estimate $V_v(\mathbf{s}_i)$ using the $m = 1, \dots, 100$ centiles of the Socio-Economic Indicator For Areas Index of Economic Resources (IER), a well established socio-economic indicator produced for SA1 areas in Australia by the ABS [46]. There are 35 SA1 areas out of 16850 that do not have an IER. For each of these, we obtained a value for the IER based on a regression of the 16815 IER values that we do have, regressed on the covariates included in our model (see Section 6.5). Finally then, we have 16850 IER values whose centiles define 100 groups of SA1 areas, as follows:

$$I_m \equiv \{i : i\text{-th SA1 area} \in m\text{-th decile}\}, \text{ for } m = 1, \dots, 100.$$

Define $q_m \equiv |I_m|$, $\bar{n}_{mj} \equiv \sum_{i \in I_m} n_{ij}/q_m$, $\bar{n}_m \equiv \sum_{i \in I_m} n_i/q_m$, and hence we set

$$V_v(\mathbf{s}_i) \equiv \sum_{j=1}^9 \frac{(\bar{n}_{mj}/\bar{n}_m)\sigma_j^2}{n_i}, \text{ for } i \in I_m, m = 1, \dots, 100. \quad (42)$$

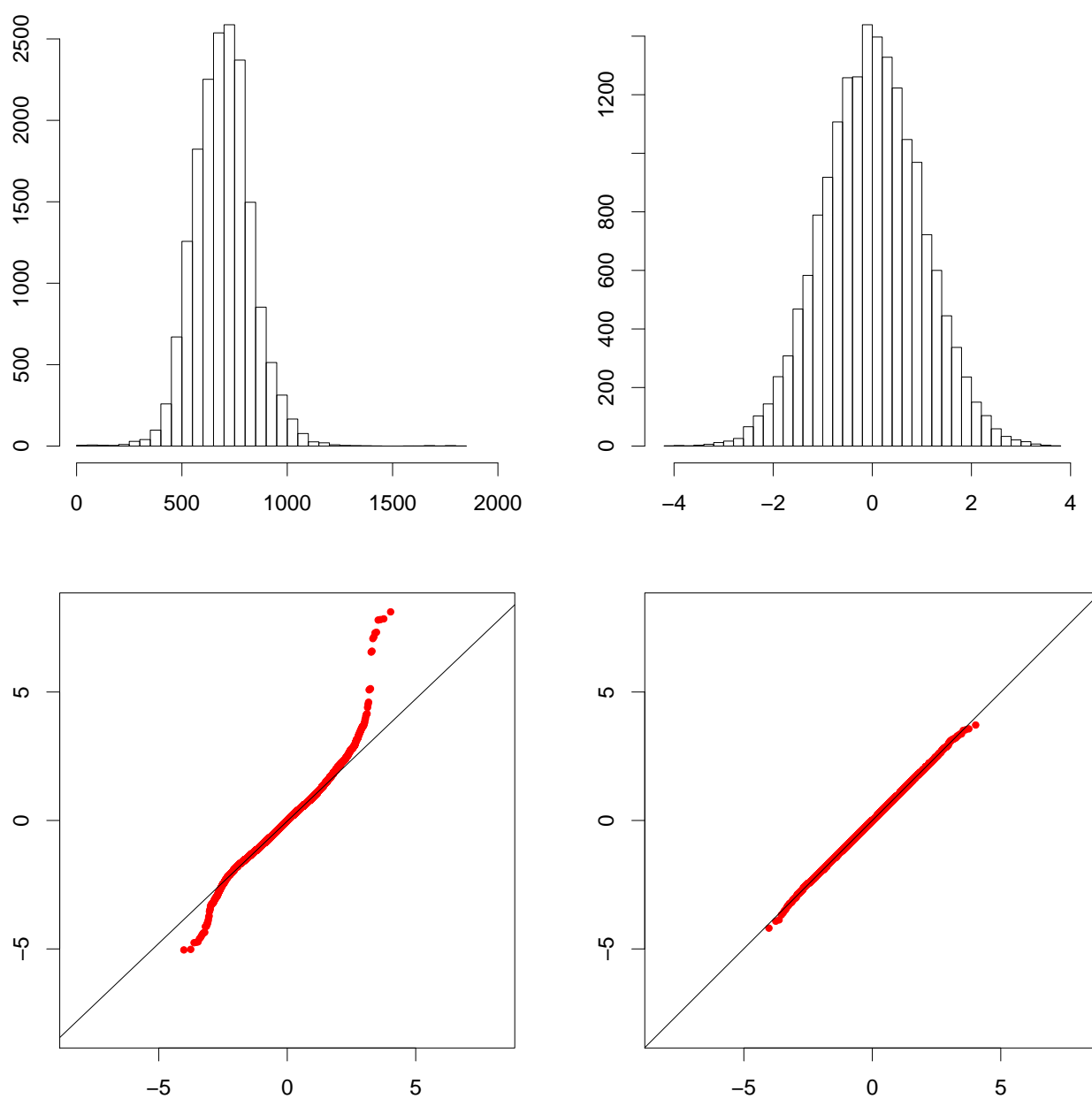
That is, in (42) we replace the *random* weights $\{(n_{ij}/n_i) : j = 1, \dots, 9\}$ given in (40) with centile-smoothed weights. This leaves a single variance, σ_v^2 , to be estimated.

6.3. Measurement Error in MWI

Known sources of measurement error in the Census are due to non-response, invalid responses, and random error purposely applied to the SA1 areas that have small populations (for confidentiality reasons; see [44] for further details). In the 2011 Census, item non-response for mean usual weekly income at the person level was approximately 7.9% [47].

Here we model the non-response as “missing completely at random.” Then for the i -th area, the observed n_{ij} is the sum of Bernoulli random variables, $n_{ij} = \sum_{k=1}^{N_{ij}} J_{ijk}; j = 1, \dots, 9$, where J_{ijk} is an indicator variable with $\Pr(J_{ijk} = 1) = p = 0.921$, obtained from the item non-response, and N_{ij} is the true number

Figure 2. Histogram and Q-Q plot of Z for 16850 SA1 areas (left panels). Histogram and Q-Q plot of standardised values (41) for the same SA1 areas (right panels).



of people. From the randomness associated with $\{J_{ijk} : k = 1, \dots, N_{ij}\}$, we calculate $\text{var}(\bar{Z}(\mathbf{s}_i)|\{Z_{ijk}\})$ using the delta method. That is,

$$\begin{aligned} \sigma_\varepsilon^2 V_\varepsilon(\mathbf{s}_i) &= E_{\mathbf{n}}(\text{var}(\bar{Z}(\mathbf{s}_i)|\{Z_{ijk}\})) + \text{var}_{\mathbf{n}}(E(\bar{Z}(\mathbf{s}_i)|\{Z_{ijk}\})) \\ &\simeq \left(\frac{N_i p + p - 1}{N_i(1-p + N_i p)^2} + \frac{4p(1-p)}{(1-p + N_i p)^3} - \frac{1}{N_i^2} + \frac{2(1-p)}{N_i^3 p} \right) \sum_{j=1}^9 N_{ij} \sigma_j^2; i = 1, \dots, 16850. \end{aligned} \quad (43)$$

Recall that N_{ij} is the fully enumerated population total for area i and stratum j , and we define $N_i \equiv \sum_{j=1}^9 N_{ij}$. To estimate (43), we use $n_{ij}p^{-1}$ for N_{ij} and $n_i p^{-1}$ for N_i . Because the measurement-error calculations come from a separate study of item non-response, the values given by (43) are considered as known, and hence $\sigma_\varepsilon^2 \mathbf{V}_\varepsilon \equiv \text{diag}(\sigma_\varepsilon^2 V_\varepsilon(\mathbf{s}_i) : i = 1, \dots, 16850)$ is known.

A summary statistic giving the measurement-error component of variability as a fraction of the total variability is:

$$\frac{\sigma_\varepsilon^2 \sum_{i=1}^{16850} V_\varepsilon(\mathbf{s}_i)}{\sum_{i=1}^{16850} (\bar{Z}(\mathbf{s}_i) - \bar{\bar{Z}})^2},$$

where $\bar{\bar{Z}} \equiv \sum_{i=1}^{16850} \bar{Z}(\mathbf{s}_i)/16850$. In our case, this is equal to 0.01%, which is very small.

6.4. Specifying the Spatial-Dependence Matrix

The simple exploratory model we proposed in Section 6.2 does not account for spatial dependence between the SA1 areas. Now that we have captured the heteroskedasticity (Section 6.2) and measurement error (Section 6.3), we can build a SAR model (see Section 2.1) for the MWI values, $\{Y(\mathbf{s}_i)\}$, defined for a fully enumerated population. That is, assume that $\mathbf{Y} \equiv (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_{16850}))'$ has covariance matrix given by (3), which is parameterised in terms of ρ and σ_v^2 .

Because there are $n = 16850$ small areas, fitting the SAR model is numerically challenging, so we turn to estimating the SAR parameters ρ and σ_v^2 by fitting the SAR-calibrated SRE model given in Section 4. Recall that we choose $\mathbf{U}_1 = (\mathbf{u}_1, \dots, \mathbf{u}_r)$, to be the matrix containing the n -dimensional eigenvectors corresponding to the r largest absolute eigenvalues in (23), which are obtained from a spectral decomposition of the GMO, (21), for a heteroskedastic SAR model.

The basis functions \mathbf{U}_1 , and the SRE-model parameters \mathbf{K} and σ_ξ^2 all depend on the spatial-dependence matrix \mathbf{H} in (3). In NSW, there is substantial variation in the physical areas of the SA1s and in the spatial proximity of neighbouring areas. Along the eastern coastline and in the major urban centres, SA1s are reasonably uniformly distributed, whereas in other parts of the state there is much greater asymmetry in the number and proximity of neighbouring areas.

A common choice for \mathbf{H} is the symmetric adjacency matrix:

$$\mathbf{H} = \mathbf{A} \equiv (a_{ii'}), \quad (44)$$

where $a_{ii'} = 1$ if the i -th area shares a common boundary with the i' -th area; and $a_{ii'} = 0$, otherwise. However, an asymmetric \mathbf{H} is more reasonable in this instance where distance or the shortest travel time between areas is more important than sharing a boundary. On a regular spatial lattice, the *eight nearest neighbours* are also the adjacent neighbours but, on an irregular lattice, they can be very different (see

Figure 1b). Here we model \mathbf{H} based on the eight nearest neighbours of the irregular SA1 spatial lattice; that is,

$$\mathbf{H} = \mathbf{E} \equiv (e_{ii'}), \quad (45)$$

where $e_{ii'} = I(i' \in E(i))$, for

$$E(i) \equiv \{i' : d_{ii'} \leq \text{eighth-largest distance from } i\}$$

and $d_{ii'}$ is the distance between the respective centroids of the i -th and i' -th areas,

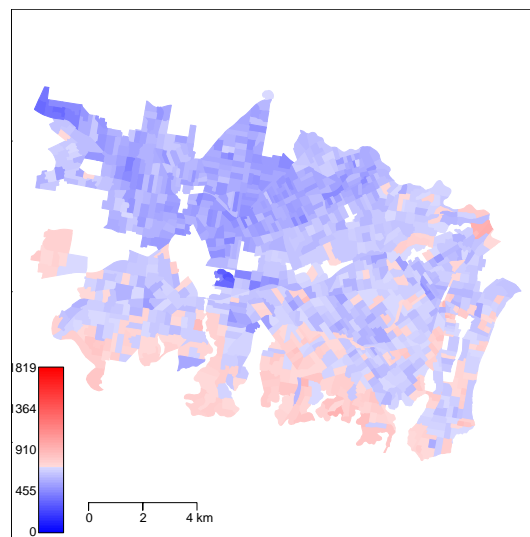
Clearly, from Figure 1b, if $i' \in E(i)$, it may not be true that $i \in E(i')$, and hence \mathbf{E} will generally be asymmetric. In our analysis, we fitted the SAR-calibrated SRE model using \mathbf{H} equal to both (row-standardised) \mathbf{A} and (row-standardised) \mathbf{E} , and we compared them.

6.5. Statistical Inference

In this section, we obtain parameter estimates for ρ , σ_v^2 , and β from the SAR-calibrated SRE model, using the methodology outlined in Section 5. Initially, we estimate the parameters for a small subset of the SA1 data obtained from the 1230 SA1 areas that are located within the 2011 Census SA4 region of South-Western (SW) Sydney (Figure 3). We use these results to help select the number of basis functions, r , to use when modelling the full NSW dataset. With $n = 1230$, the SAR-model estimate $\hat{\rho}_{SAR}$ for ρ can be computed, and hence r is chosen so that $\hat{\rho}$ (which depends on r) is close to the target value $\hat{\rho}_{SAR}$.

We subsequently analyse the full NSW dataset using this selected value of r in a SAR-calibrated SRE model. Clearly, direct likelihood-based parameter estimates based on the SAR model are no longer possible for all of NSW, due to the size of the dataset ($n = 16850$).

Figure 3. SA1 level imputed MWI for SW Sydney.



While the basic heteroskedastic model for MWI in Section 6.2 fitted well, we cannot use it for inference because its mean depends on the random counts $\{n_{ij}\}$ used to impute the data. Socio-economic and demographic covariates are generally used in an attempt to explain the variability in MWI. Using various selection criteria, we settled on the following explanatory variables: the proportion of males, out

of the total population, in the age groups 15-24, 25-64, and ≥ 65 (3 covariates); the proportion of females, out of the total population, in the same age groups (3 covariates); the median mortgage repayment, the median rent, the average household size, and the average number of persons per bedroom (4 covariates); and the overall mean effect (1 covariate). That is, we fitted a regression, $\mathbf{X}\boldsymbol{\beta}$, for the mean of the spatial model, where the number of regression parameters was $p = 11$. The estimate of $\boldsymbol{\beta}$ is important and interpretable, and it may even be used by governments to make policy.

We specify the matrices that we use to represent heteroskedasticity, measurement error, and spatial dependence in the model using (42), (43), and (44) or (45) from Sections 6.2, 6.3, and 6.4, respectively. To complete our specification of the model, we use the term $\sigma_{\xi}^2 \mathbf{V}_{\xi}$ to capture the additional variance not captured by the spatial basis functions included in the model. We expect the variance to depend on the number of observations in each area, and hence we set $V_{\xi}(\mathbf{s}_i) = c/n_i; i = 1, \dots, 16850$, where $c = \text{trace}(\mathbf{V}_v)/(\sum_{i=1}^{16850} 1/n_i)$ is a scaling constant that makes σ_{ξ}^2 comparable to σ_v^2 .

We use the restricted-maximum-likelihood-estimation method described in Section 5 to obtain REML estimates, $\hat{\rho}$ and $\hat{\sigma}_v^2$, for the SW Sydney and the NSW datasets, for the spatial-dependence matrices \mathbf{A} and \mathbf{E} defined by (44) and (45), respectively. In each case, we used a grid-search algorithm to obtain $\hat{\rho}$. For each value of ρ , we obtained a starting value, $\sigma_{v,0}^2(\rho)$ using (34) with $\mathbf{N} = \text{diag}(n_i : i = 1, \dots, 16850)$, and then we iteratively estimated $\hat{\sigma}_v^2(\rho)$. We calculated the profile restricted log-likelihood for each value of $(\rho, \hat{\sigma}_v^2(\rho))$, from which we obtained $\hat{\rho} \equiv \arg \min(\mathcal{L}_R^*(\rho))$, and hence $\hat{\sigma}_v^2 \equiv \hat{\sigma}_v^2(\hat{\rho})$. Finally, we used (37) to obtain generalised-least-squares parameter estimates, $\hat{\boldsymbol{\beta}}$.

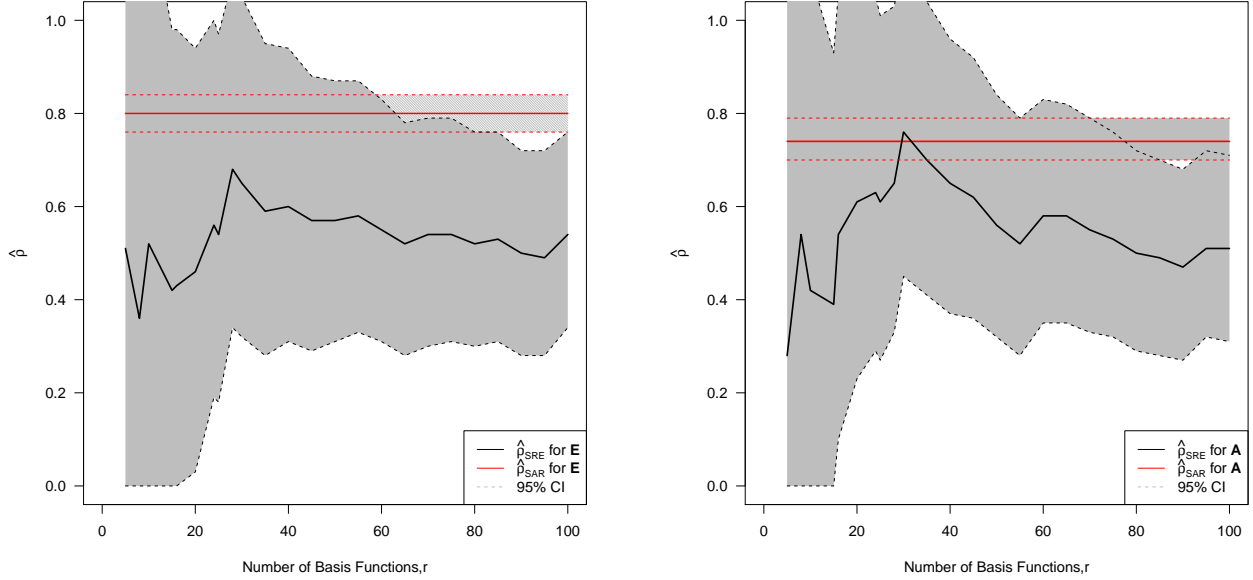
For the SW Sydney dataset, we compared $\hat{\rho}$ and $\hat{\sigma}_v^2$ with the target SAR-model estimates. Hence, we use restricted maximum likelihood estimation and account for the measurement error to obtain SAR-model parameter estimates, $\hat{\rho}_{SAR}$ and $\hat{\sigma}_{v,SAR}^2$.

Based on Figure 4 for SW Sydney, we see that for $r \geq 25$ spatial basis functions, the SAR-calibrated SRE model yields a 95% confidence interval for ρ that contains $\hat{\rho}_{SAR} = 0.8$ for \mathbf{E} and $\hat{\rho}_{SAR} = 0.74$ for \mathbf{A} . From this and plots of SAR-calibrated estimates of ρ and σ_v^2 versus r for all of NSW, we chose $r = 25$ basis functions for our dimension-reduced analysis of MWI in the 16850 SA1 areas of NSW shown in Figure 1. To choose which spatial-dependence matrix to use for the NSW dataset, we compared the magnitude of the restricted log-likelihood functions for SW Sydney obtained using \mathbf{A} and using \mathbf{E} , since the number of parameters is the same in both models. The asymmetric spatial-dependence matrix \mathbf{E} resulted in a smaller value for the negative restricted log-likelihood, so we used \mathbf{E} to fit the model for NSW.

We fit the resulting SAR-calibrated SRE model given by (31) and (30) to the full NSW dataset, to obtain SAR-calibrated SRE-model REML estimates, $\hat{\rho} = 0.74$ and $\hat{\sigma}_v^2(\hat{\rho}) = (87)^2$. We also obtain $\hat{\zeta} \equiv \zeta(\hat{\rho}) = 0.87$, and hence $\hat{\sigma}_{\xi}^2 = (100.3)^2$. From these results, we conclude that MWI exhibits substantial spatial dependence in NSW, after allowing for the covariates included in the regression model.

Table 1 compares the SAR-calibrated SRE-model estimates of $\boldsymbol{\beta}$ given by (37) with the ordinary-least-squares (OLS) regression parameter estimates, $\hat{\boldsymbol{\beta}}_{OLS} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$; the latter estimates ignore heteroskedasticity and spatial dependence. For *all* of the covariates except the intercept and “Female, 15-24 years”, the OLS estimates do not fall within the 95% CI for their respective $\boldsymbol{\beta}$ s obtained using the SAR-calibrated SRE model. The OLS estimate $\hat{\boldsymbol{\beta}}_{OLS}$ is based on the overly simple model,

Figure 4. SAR-calibrated SRE-model REML estimates for $\hat{\rho}$ and $\hat{\sigma}_v^2$ for SW Sydney as a function of r for **E** (left hand panel) and for **A** (right hand panel); the SAR model estimate $\hat{\rho}_{SAR}$ and 95% profile log-likelihood confidence intervals (CI) for ρ are included for comparison. (36).



$\mathbf{Z} \sim \text{Gau}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, and for this dataset it is inadequate. This demonstrates the importance of capturing small-scale heteroskedastic spatial variation in our analysis of MWI.

Table 1. MWI for NSW. Ordinary-least-squares regression-parameter estimates, generalised-least-squares regression-parameter estimates obtained from fitting a SAR-calibrated SRE-model, and 95% confidence intervals (CIs). The SRE model is based on 25 basis functions and $\mathbf{H} = \mathbf{E}$ given by (45).

	β_0	β_{15-24}^{Male}	β_{25-64}^{Male}	$\beta_{\geq 65}^{Male}$	$\beta_{15-24}^{Fem.}$	$\beta_{25-64}^{Fem.}$
$\hat{\boldsymbol{\beta}}_{SRE}$	463	47.5	773	-782	-201	675
CI_{SRE}	(432,494)	(-6.7,102)	(732,814)	(-836,-729)	(-256,-146)	(622,728)
$\tilde{\boldsymbol{\beta}}_{OLS}$	481	-95.6	719	-695	-177	789
	$\beta_{\geq 65}^{Fem.}$	$\beta_{Mort.}$	β_{Rent}	β_{Size}	β_{Beds}	
$\hat{\boldsymbol{\beta}}_{SRE}$	346	0.0443	0.298	-234	-15.1	
CI_{SRE}	(302,390)	(0.042,0.0466)	(0.287,0.309)	(-244,-224)	(-18.7,-11.3)	
$\tilde{\boldsymbol{\beta}}_{OLS}$	219	0.0472	0.197	-245	-7.84	

7. Discussion

This paper has considered the specification and estimation of a hierarchical model that includes a computationally efficient Spatial Random Effects (SRE) model for a latent spatial process and an additive measurement-error model for the observed data. Using the Generalised Moran operator, we calibrate

this model to a SAR model, a model that is widely used in spatial econometrics but is computationally prohibitive for large datasets. We use restricted maximum likelihood on a computationally efficient SRE model to estimate the SAR model's covariance parameters.

Implementation of our model is demonstrated using Australian Census data for mean income. It is clear from the SAR-calibrated SRE model fitted to NSW's 16850 SA1 areas that a regression model with the 11 covariates and independent residuals does not properly capture the small-scale variability. This (heteroskedastic) spatial variability is important, which is reflected in the estimate for ρ , namely $\hat{\rho} = 0.74$ for 25 basis functions. It is also reflected in the change of regression-coefficient estimates when the spatial variability is used to improve both the estimates and their estimation variance.

Notice the role of model calibration here, which is different from model fitting. The SRE model is calibrated to be close to the SAR model of interest, and then the computationally efficient SAR-calibrated SRE model is fitted. Importantly, the parameters that are fitted are those of the original, interpretable SAR model. This enables SAR model parameters to be estimated for much larger datasets than is currently possible.

It is easy to see that our SAR-calibrated SRE model can be adapted to obtain a CAR-calibrated SRE model. In some sense, this represents the converse of the calibration undertaken by Rue and Tjelmeland [48], who find the Gaussian Markov random field (GMRF) that comes closest to a geostatistical model; note that the CAR model is a GMRF and the SRE model is a geostatistical model.

Besides the advantages of the SAR-calibrated SRE model in producing parameter estimates, the r spatial basis functions in \mathbf{S} could be used to identify hidden, possibly overlapping, regions or classes of areas. The associated random effects can be easily predicted, and hence regions or classes having higher or lower values of MWI could also be predicted. Moreover, as the covariance matrix \mathbf{K} of the random effects is not constrained to be diagonal, the SRE model captures additional dependence associated with these classes.

Acknowledgements

This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program; and it was partially supported by a 2015-2017 Australian Research Council Discovery Project.

References

1. Whittle, P. On stationary processes in the plane. *Biometrika* **1954**, *41*, 434–449.
2. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer Academic Publishers: Dordrecht, NL and Boston, MA, 1988.
3. Pace, R.K.; Barry, R. Sparse spatial autoregressions. *Statistics and Probability Letters* **1997**, *33*, 291–297.
4. Smirnov, O.A. Computation of the information matrix for models with spatial interaction on a lattice. *Journal of Computational and Graphical Statistics* **2005**, *14*, 910–927.
5. Pace, R.K.; Le Sage, J.P. Chebyshev approximation of log-determinants of spatial weight matrices. *Computational Statistics and Data Analysis* **2004**, *45*, 179–196.
6. Wikle, C.K. Low rank representations as models for spatial processes. In *Handbook of Spatial Statistics*; Gelfand, A.E.; Diggle, P.J.; Fuentes, M.; Guttorp, P., Eds.; Chapman and Hall/CRC Press: Boca Raton, FL, 2010; pp. 107–118.
7. Lopes, H.F.; Gamerman, D.; Salazar, E. Generalized spatial dynamic factor models. *Computational Statistics and Data Analysis* **2011**, *55*, 1319–1330.
8. Banerjee, S.; Gelfand, A.E.; Finley, A.O.; Sang, H. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B* **2008**, *70*, 825–848.
9. Finley, A.O.; Sang, H.; Banerjee, S.; Gelfand, A.E. Improving the performance of predictive process modelling for large datasets. *Computational Statistics and Data Analysis* **2009**, *53*, 2873–2884.
10. Eidsvik, J.; Finley, A.O.; Banerjee, S.; Rue, H. Approximate Bayesian inference for large spatial datasets using predictive process models. *Computational Statistics and Data Analysis* **2012**, *56*, 1362–1380.
11. Cressie, N.; Johannesson, G. Spatial prediction for massive data sets. Mastering the data explosion in the earth and environmental sciences: proceedings of the Australian Academy of Science Elizabeth and Frederick White Conference, 2006, pp. 1–11.
12. Cressie, N.; Johannesson, G. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B* **2008**, *70*, 209–226.
13. Kelejian, H.H.; Prucha, I.R. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* **1999**, *40*, 509–533.
14. Hughes, J.; Haran, M. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society, Series B* **2013**, *75*, 139–159.
15. Sengupta, A.; Cressie, N. Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions. *Spatial Statistics* **2013**, *4*, 14–44.
16. Shi, T.; Cressie, N. Global statistical analysis of MISR aerosol data: a massive data product from NASA's Terra satellite. *Environmetrics* **2007**, *18*, 665–680.
17. Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; Wiley: Hoboken, NJ, 2011.
18. Royle, J.A.; Wikle, C.K. Efficient statistical mapping of avian count data. *Environmental and Ecological Statistics* **2005**, *12*, 225–243.

19. Buhmann, M.D. *Radial Basis Functions: Theory and Implementations*; Cambridge University Press: Cambridge and New York, 2003.
20. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer, 2001.
21. Katzfuss, M.; Cressie, N. Bayesian hierarchical spatio-temporal smoothing for very large datasets. *Environmetrics* **2012**, *23*, 94–107.
22. Nychka, D.; Wikle, C.K.; Royle, J.A. Multiresolution models for non-stationary spatial covariance functions. *Statistical Modelling* **2002**, *2*, 315–331.
23. Bradley, J.R.; Cressie, N.; Shi, T. Selection of rank and basis functions in the spatial random effects model. In *2011 Proceedings of the Joint Statistical Meetings*; American Statistical Association: Alexandria, VA, 2011; pp. 3393–3406.
24. Griffith, D.A. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra and its Applications* **2000**, *321*, 95–112.
25. Griffith, D.A. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A* **2008**, *40*, 2751–2769.
26. Tiefelsdorf, M.; Griffith, D.A. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A* **2007**, *39*, 1193–1221.
27. Reich, B.J.; Hodges, J.S.; Zadnik, V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **2006**, *62*, 1197–1206.
28. Bradley, J.R.; Holan, S.H.; Wikle, C.K. Mixed effects modelling for areal data that exhibit multivariate spatio-temporal dependencies **2014**. *arXiv preprint: arXiv:1407.7479v2 [stat.ME]*.
29. Besag, J.; York, J.; Mollié, A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **1991**, *43*, 1–20.
30. Cliff, A.D.; Ord, J.K. *Spatial Processes: Models and Applications*; Pion: London, UK, 1981.
31. Cressie, N. *Statistics for Spatial Data*, 2nd ed.; John Wiley and Sons, Inc: New York, 1993.
32. Le Sage, J.P. Bayesian estimation of spatial autoregressive models. *International Regional Science Review* **1997**, *20*, 113–129.
33. Anselin, L. Under the hood. *Agricultural Economics* **2002**, *27*, 247–267.
34. LeSage, J.P.; Pace, R.K. *Introduction to Spatial Econometrics*; CRC Press: Boca Raton, FL, 2009.
35. Sengupta, A.; Cressie, N. Empirical hierarchical modelling for count data using the spatial random effects model. *Spatial Economic Analysis* **2013**, *8*, 389–418.
36. Li, H.; Calder, C.A.; Cressie, N. One-step estimation of spatial dependence parameters: Properties and extensions of the APLE statistic. *Journal of Multivariate Analysis* **2011**, *105*, 68–84.
37. Cressie, N.; Lahiri, S.N. The asymptotic distribution of REML estimators. *Journal of Multivariate Analysis* **1993**, *45*, 217–233.
38. Schabenberger, O.; Gotway, C.A. *Statistical Methods for Spatial Data Analysis*; Chapman & Hall/CRC: Boca Raton, FL, 2005.

39. Tiefelsdorf, M. *Modelling Spatial Processes: The identification and analysis of spatial relationships in regression residuals by means of Moran's I*; Vol. 87, *Lecture notes in earth sciences*, Springer-Verlag: Berlin and New York, 2000.
40. Tiefelsdorf, M.; Boots, B. The exact distribution of Moran's I. *Environment and Planning A* **1995**, *27*, 985–999.
41. Griffith, D.A. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer* **1996**, *40*, 351–367.
42. Higham, N.J. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications* **1988**, *103*, 103–118.
43. Henderson, H. V.; Searle, S. R.. On deriving the inverse of a sum of matrices. *SIAM Review* **1981**, *23*, 53–60.
44. Australian Bureau of Statistics. 2011 Australian Census of Population and Housing: New South Wales (SA1 and SA2 Level areas), Basic Community Profile, Table 17, 2011.
45. Australian Bureau of Statistics. Survey of Income and Housing 2011-12, Confidentialised Unit Record File, 2012.
46. Australian Bureau of Statistics. Socio-Economic Indexes for Areas (SEIFA) 2011 – Technical Paper: ABS Catalogue no. 2033.0.55.001, 2013.
47. Australian Bureau of Statistics. 2011 Australian Census of Population and Housing: Data Quality Statement for Total Personal Income (weekly) (INCP), 2011.
48. Rue, H.; Tjelmeland, H. Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics* **2002**, *29*, 31–49.